

correction command and corrected text in the form of a pronunciation of a word to be corrected.

Gould simply does not describe or suggest processing an utterance that includes both a correction command and corrected text in the form of a pronunciation of a word to be corrected. As noted by the Examiner, Gould's potential correction commands are "Choose-N" and "Scratch-That". Clearly, neither of these commands, standing alone, includes corrected text in the form of a pronunciation of a word to be corrected. Apparently recognizing this shortcoming of Gould, the Examiner attempts to define the term "utterance" overbroadly, so as to include, for example, utterance of the word "very" (i.e., a first utterance), followed by the later utterance of the command "Choose-3" (i.e., a second utterance), within the confines of a single utterance.

This treatment of the two utterances as a single utterance contradicts both the definition of an utterance set forth in the application, and the well-understood definition of an utterance as a term in the art, as evidenced by Gould. The application at, for example, pages 5-6, notes that utterances are "separated from one another by a pause having a sufficiently large predetermined duration (e.g., 160-250 milliseconds)" and that "[e]ach utterance may include one or more words of the user's speech." Gould defines an utterance similarly: "the utterance to be recognized will normally be proceeded [sic, preceded] and followed by silence" (col. 2, lines 3-5).

Indeed, in the example noted by the Examiner, Gould clearly treats "very" as an utterance separate from the utterance of "Choose-3". As discussed by Gould at col. 7, line 64 to col. 9, line 25, and illustrated in Fig. 5, an utterance may be a word, which results in removal of the previous choice window, simulated typing of the best scoring word, and creating of a new choice window (see steps 218-224 of Fig. 5). An utterance also may be a choice command, which results in replacement of the best scoring word with the chosen word and removal of the choice window (see steps 216 and 226-234 of Fig. 5). As noted by Gould at col. 27, lines 11-15 and illustrated in Fig. 46, saying "very" causes "vary" to be displayed along with a choice window, thus establishing "very" as a first utterance (i.e., an utterance of a word). When the user later says "Choose-3", this is treated as a second, separate utterance (i.e., an utterance of a choice command). See col. 27, lines 41-46.

Moreover, even if "very" and "Choose-3" could somehow be said to constitute parts of a single utterance, "very" could not be said to constitute "corrected text," since, at the time "very" is spoken, there is no error to be corrected. Rather, the misrecognition of "very" as "vary" is the error that requires correction through use of the "Choose-3" command. For each of these reasons, claim 1 is not anticipated by Gould's "Choose-N" command.

The Examiner refers to the "Scratch-That" command as somehow identifying corrected text by stating that "identification and removal of incorrect text corrects text in some cases." This is incorrect for at least the following reasons. First, the process of correcting text, as embodied in the "Scratch-That" command, and pointed to by the Examiner, is not equivalent to identifying corrected text, as recited in claim 1. Identifying corrected text is used as a way of correcting text. However, the process of correcting text does not require the identifying of corrected text. Rather, as evidenced by the "Scratch-That" command, the process of correcting text may simply include the removal of incorrectly recognized text.

The Examiner also indicates that Gould's adaptive training subroutine somehow indicates that the portion of the recognition result for an utterance of a correction command includes a pronunciation of a word to be corrected, as recited in claim 1. However, the adaptive training subroutine (as described in Fig. 12 and called in the routine of Fig. 5 of Gould) is only used to improve word models for a vocabulary. See Gould at col. 10, lines 3-16. The Examiner states at page 10 of the office action that the last token recognized is automatically stored for the entry in the OOPS buffer to correct the word if the word turns out to have been misrecognized. However, the stored token is never used by the commands in Gould to correct the word. The stored token is used initially by the system to determine the

words in the choice list. Then, the stored token is used after a word has been correctly labeled (using, for example, the CHOOSE-N command) to perform adaptive training on word models. Because the token in Gould's system is used in the adaptive training subroutine only after the user has selected the correct word using a command such as CHOOSE-N, the token cannot be considered to be corrected text including a "pronunciation of a word to be corrected," as recited in claim 1.

For these reasons, Applicants submit that Gould in no way describes or suggests the subject matter of claim 1. Claims 2-6 and 12 depend from claim 1 and are allowable for the reasons set forth above, and for containing allowable subject matter in their own right. Accordingly, Applicants request withdrawal of the rejection of claims 1-6 and 12.

Claims 8-11 and 13-24, all of which depend from claim 1, stand rejected as being obvious over Gould in view of Roberts (US Patent No. 5,027,406). Roberts, however, fails to cure the deficiencies of Gould. As was discussed in the interview granted on 1/14/99 and as was conceded by the Examiner, the correction commands ("start\_comletter" and "backspace") of Roberts do not comprise a pronunciation of a word to be corrected, as recited in claim 1. Therefore, since both Gould and Roberts lack this feature, any possible combination of Gould and Roberts would fail to describe or suggest the combination of features of claim 1.

For this reason, Applicants also request withdrawal of the rejection of claims 8-11 and 13-24.

Claims 25 and 27-30 stand rejected as being obvious over Roberts et al. in view of Junqua (US Patent No. 5,677,990).

Independent claim 25 recites a method for recognizing a spelling of a word in computer-implemented speech recognition. The method includes performing speech recognition on an utterance to produce recognition results for the utterance and identifying a spelling command in the recognition results. The spelling command indicates that a portion of the utterance includes a spelling. The method further includes producing the spelling by searching a dictionary using the recognition results. Producing the spelling includes using confused spelling matching. In confused spelling matching, commonly-confused letters are treated as a single letter to identify the spelling corresponding to the portion of the utterance.

Applicants request reconsideration and withdrawal of this rejection because Roberts fails to describe or suggest using confused spelling matching and instead uses a phonetic alphabet to increase spelling recognition accuracy. Furthermore, there would have been no motivation to employ Junqua's confused spelling matching in Roberts' system because neither Roberts nor Junqua describes or suggests that confused spelling matching improves recognition accuracy in a speech recognition system that otherwise uses a phonetic alphabet.

The Examiner states on page 12 of the office action that Junqua's confused spelling matching makes letter commands or a phonetic alphabet unnecessary. The Examiner uses an example of a misrecognition of the word "invention" as "inversion" to give motivation for using Junqua in the system of Roberts. According to the Examiner, the Roberts' "starts\_eye..." command that corrects this misrecognition produces a much longer list of recognition candidates than the Examiner's suggested command "starts\_i...n...v...e...n...t...". The Examiner then states that Junqua suggests at col. 1, lines 50-67 that confused spelling matching would further decrease the response time by producing an even shorter list of candidates.

This is simply wrong. Since confused spelling matching permits a single letter to represent multiple letters, confused spelling matching will actually produce a longer list of candidates.

In addition to this fundamental flaw, the Examiner's argument includes a number of other problems. First, contrary to the Examiner's assertion, Junqua, at col. 1, lines 50-67, in no way suggests that confused spelling matching decreases the response time by producing a shorter list of candidates. Rather, that passage states that, although reasonable accuracy may be obtained using a fixed list such as a telephone directory to constrain spelling recognition, response time will increase if the size of the list increases: "response time increases quite

dramatically as the size of the list or dictionary increases." See Junqua at col. 1, lines 50-54. Therefore, according to Junqua, a system that does not use a large list (that is, a large fixed list or dictionary) decreases response time.

This is why Junqua breaks up the speech recognition process into steps as Junqua states at col. 1, lines 62-65 -- "To attain optimally short response time, the processes are performed first without costly constraints and thereafter with costly constraints, if needed, after the number of word candidates is low." Thus, confused spelling is used only after the first pass produces an N-best candidate list. See Junqua at Fig. 1, step 36, and col. 6, lines 39-49. Junqua does this because confused spelling matching, when used on an entire name dictionary, tends to increase response time. Therefore, confused spelling matching is performed on a "dynamic grammar" as opposed to the entire name dictionary. See Junqua at col. 2, lines 25-29 and Fig. 1.

Furthermore, Junqua's speech recognition system is said to gain the highest recognition accuracy not from confused spelling matching (as the Examiner suggests), but from the highly constrained recognition at the fourth pass. See Junqua at col. 8, lines 41-63.

Second, contrary to the Examiner's assertion, Fig. 15 of Roberts has nothing to do with speaking a "starts\_invent" command. Fig. 15 of Roberts does not relate to speech recognition of letters, but instead relates to processing of

typed letters. Specifically, Fig. 15 relates to a condition in which a user first types the letter "i" into the system, thus restricting the candidate list to words that begin with the letter "i". Then, the user types "n", "v", and "e". See Roberts at col. 24, lines 32-38 and col. 24, lines 54-63.

Furthermore, although the Examiner claims there is motivation for using confused spelling matching in the system of Roberts because confused spelling matching somehow improves recognition accuracy, neither Junqua nor Roberts states or implies that using confused spelling matching would improve recognition accuracy in a speech recognition system that would otherwise use a phonetic alphabet (as used in Roberts). To the contrary, Junqua implies in col. 1, lines 42-49 that using a phonetic alphabet actually eliminates the need for confused spelling matching because using a phonetic alphabet actually improves recognition accuracy: "Recognition of spoken letters is even difficult for humans ... This is why radio telephone operators are trained to use a phonetic alphabet, A-Alpha, B-Baker, C-Charlie, etc., when communicating over a noisy channel." Similarly, Roberts states at col. 19, line 57 - col. 20, line 19 that the phonetic alphabet commands are used with a restricted vocabulary (or list) in the EDITMODE. This is done in Roberts because a large vocabulary or list is not needed when using the phonetic alphabet because the phonetic alphabet produces increased recognition accuracy.

Roberts fails to describe or suggest using confused spelling matching to identify the spelling corresponding to the portion of the utterance. Moreover, for the reasons noted above, there would have been no reason to employ the confused spelling matching system of Junqua in Roberts' system. As noted above, if a special phonetic alphabet is used (such as the communications or phonetic alphabet of Roberts), spelling recognition accuracy in a speech recognizer is greatly improved. With this improved recognition accuracy, there is no need to employ confused spelling matching in Roberts' speech recognition system. Accordingly, one of ordinary skill in the art would have had no motivation to combine Roberts and Junqua in the manner suggested by the Examiner.

Claims 27-30 depend from claim 25 and are allowable for the reasons set forth above, and for containing allowable subject matter in their own right. For these reasons, Applicants request withdrawal of the rejection of claims 25 and 27-30.